

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:46 ; Search time 170.72 Seconds

(Without alignments)
23.864 Million cell updates/sec

Title: US-09-331-631a-21_COPY_32_91

Perfect score: 343

Sequence: 1 TENPCAQRCLSCQOEPPDL.....DTGATNRHPPERTGRGP 60

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98.5	28.7	639	B24810	beta-conglycinin a
2	93	27.1	605	FMSYBA	beta-conglycinin a
3	93	27.1	605	S20007	beta-conglycinin a
4	75	21.9	411	T29475	hypothetical prote
5	70	20.4	221	T31863	hypothetical prote
6	69	20.1	1187	T18355	hypothetical prote
7	68.5	20.0	1513	T23681	hypothetical prote
8	67.5	19.7	438	T31889	hypothetical prote
9	67.5	19.7	445	T31888	hypothetical prote
10	66.5	19.4	509	S08059	alpha-globulin typ
11	66.5	19.4	771	D49423	semaphorin III pre
12	66	19.2	422	T29699	hypothetical prote
13	66	19.2	388	T31887	hypothetical prote
14	66	19.2	388	T31888	hypothetical prote
15	66	19.2	605	S06398	alpha-globulin typ
16	64	18.7	1300	A36502	insulin receptor-r
17	63.5	18.5	588	FWCNAB	alpha-globulin B p
18	63.5	18.5	666	I58169	semaphorin III - m
19	63.5	18.5	772	I48747	semaphorin D - mou
20	63	18.4	572	T29880	hypothetical prote
21	63	18.4	657	S32991	hypothetical prote
22	62.5	18.2	703	G69186	ribonucleoside-dip
23	61.5	17.9	894	FN0667	peroxinectin-like
24	61	17.8	151	S22145	hypothetical prote
25	61	17.8	181	T03720	4-comarate--COA 1
26	61	17.8	545	A39827	4-comarate--COA 1
27	61	17.8	728	B39827	4-comarate--COA 1
28	61	17.8	728	T20561	hypothetical prote
29	60.5	17.6	263	T27641	hypothetical prote

30	60.5	17.6	375	2	T16248	hypothetical prote
31	60.5	17.6	4391	2	A38096	perlecan precursor
32	60	17.5	419	2	S69207	vascular endotheli
33	60	17.5	699	2	S27768	RNA-directed DNA p
34	59.5	17.3	772	2	A49069	collapsin - chicke
35	59	17.2	186	2	A43910	ultra-high-sulfur
36	59	17.2	294	2	T23682	hypothetical prote
37	59	17.2	333	2	B45558	epidermal growth f
38	59	17.2	342	2	C45558	epidermal growth f
39	59	17.2	356	2	A25918	thrombomodulin - b
40	59	17.2	366	2	D45558	epidermal growth f
41	59	17.2	634	2	T30136	hypothetical prote
42	59	17.2	860	2	T16892	hypothetical prote
43	59	17.2	1717	1	A45558	epidermal growth f
44	58.5	17.1	330	2	T25169	hypothetical prote
45	58.5	17.1	342	2	T46909	hypothetical prote

ALIGNMENTS

```

RESULT 1
B24810
beta-conglycinin alpha chain precursor - soybean
C:Species: Glycine max (soybean)
C>Date: 07-Mar-1988 #sequence_revision 30-Jun-1991 #text_change 20-Aug-1999
C:Accession: B24810; S16337
R:Doyle, J.J.; Schuller, M.A.; Godette, W.D.; Zenger, V.; Beachy, R.N.; J. Biol. Chem. 261, 9228-9238, 1986
A:Title: The glycosylated seed storage proteins of Glycine max and Phaseolus vulgaris
A:Reference number: A24810; MUID:86250867
A:Accession: B24810
A:Molecule type: DNA
A:Residues: 1-639 <DOY>
A:Cross-references: GB:M13759; NID:g169928; PID:MAB01374.1; PID:g169929
A>Note: The authors translated the codon GGT for residue 352 as Glu
R:Schuller, M.A.; Schmitt, E.S.; Beachy, R.N.
Nucleic Acids Res. 10, 8225-8244, 1982
A:Title: Closely related families of genes code for the alpha and alpha' subunits of
A:Reference number: S16337; MUID:6314288
A:Accession: S16337
A:Molecule type: DNA
A:Residues: 361-639 <SCH>
C:Genetics:
A:Introns: 278/1; 355/3; 382/3; 481/3; 575/1
C:Superfamily: glycinin
C:Keywords: seed; storage protein

Query Match 28.7%; Score 98.5; DB 2; Length 639;
Best Local Similarity 30.6%; Pred. No. 0.0029;
Matches 22; Conservative 10; Mismatches 25; Indels 15; Gaps 3;

OY 2 ENPCAQRCLSCQOEPPDLKQACESECT--KLEIDPCYVDTG-----ATQR 48
DB 31 QNPHNKCLRCSEKSYRNQACHARCMLKVEEBC--EEQIPRRPRQHPREERQ 88
OY 49 HPGFRRGRGP 60
DB 89 HGEKEDEGEQP 100

RESULT 2
FMSYBA
beta-conglycinin alpha chain precursor - soybean
C:Species: Glycine max (soybean)
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C:Accession: S14681; S74124; S06714
R:Sebastiani, F.V.; Farrell, L.B.; Schuller, M.A.; Beachy, R.N.
Plant Mol. Biol. 15, 197-201, 1990
A:Title: Complete sequence of a cDNA of alpha subunit of soybean beta-conglycinin.
A:Reference number: S14681; MUID:91355860
A:Accession: S14681

```

A:Molecule type: mRNA
A:Residues: 1-605 <SEB>
A:Cross-references: EMBL:X17698; NTD:g18535; PDB:CA53691.1; PDB:g18536
R:Shuto, A.D.; Kakhovskaya, I.A.; Baskirgina, A.S.; Bulmaga, V.P.; Horstmann, C.; Muenh
Eur. J. Biochem. 241, 221-228, 1996
A>Title: Limited proteolysis of beta-conglycinin and glycinin, the 7S and 11S storage g
A:Reference number: S74123; MUID:97054613
A:Accession: S74124
A:Molecule type: protein
A:Residues: 188-196, 'H', 198, 'N', 200, 'X', 202-203, 397-408, 'X', 410, 'X', 412-417, 'X', 419-420.
A:Experimental source: seed
C:Superfamily: glycinin
C:Keywords: glycoprotein; seed; storage protein
F:1-22/Domain: signal sequence #status predicted <IG>
F:23-62/Domain: propeptide #status predicted <PRO>
F:63-605/Product: beta-conglycinin alpha chain #status predicted <MAT>
F:261, 517/Blinding site: carbohydrate (Asn) (covalent) #status predicted

Query Match	27.1%	Score 93	DB 1	Length 605
Best Local Similarity	31.4%	Pred. No.	0.011	
Matches	22	Conservative	8	Mismatches 16; Indels 24; Gaps 3;

```

QY      2 ENPCARCLQSQQDEPDDLKQKACESRCYLEYD-----PRCVYDTGATNQRRH 50
      III :IIII I I : :II :II I :
Db      31 ENPKHNKCLQSCNSESRDSYRNQACHARNCLKVEKECECEGETPRPR-----PRQHP 83

```

```
Qy      51 -----PGER 54
          ||:
Db      84 EREPOQGEK 93
```

RESULT 3
S20007
beta-conglycinin alpha chain precursor - soybean
C:Species: Glycine max (soybean)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Oct-1997
C:Accession: S20007
R:Reliever, J.M.; Dickinson, C.D.; Dickinson, L.A.; Nielsen, N.C.
Plant Mol. Biol. 18, 259-274, 1992
Title: Synthesis and assembly of soybean beta-conglycinin in vitro.
A:Reference number: S20007; MID:92119248

A;Status: preliminary
A:Molecule type: mRNA
A;Residues: 1-605 <LEL>
C;Superfamily: glycinin

Query Match	27.1%	Score	93	DB	2	Length	605
Best Local Similarity	31.4%	Pred. No.	0.011				
Matches	22	Conservative	8	Mismatches	16	Indels	24
						Gaps	3

```

0Y      2 ENPCACORCLOSCQOEPPDLKQKACESRCTKLEYD-----PRCVYDTGATNQRRH 50
      III :IIIIII I I : :II :II I : : II
Db     31 ENPKHKKCLQSCNSERSDYRNQACHARNLLKVEKEECEEGEIPRPR-----PRQPHR 83

```

```

Oy  51  -----PGER 54
      III:
Db   84  EREPOOGEK 93

```

hypothetical protein T01D1.6 - *Caenorhabditis elegans*

C:\Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:\Accession: T29475

Submitted to the EMBL Data Library, November 1996
A;Description: The sequence of C. elegans cosmid T01D1.

A;Reference number: Z20623
A;Accession: T29475

A:Status: preliminary; translated from CB/EMBL/DDbj
A:Molecule type: DNA
A:Residues: 1-411 <RRA>
A:Cross-references: EMBL:U80455; PIDN:AMB37867.1; GSPDB:GN00020; CESP:T01D1.6
A:Experimental source: Strain Bristol N2; clone T01D1
C:Genetics:
A:Gene: CESP:T01D1.6
A:Map position: 2
A:Introns: 25/3; 304/3
C:Superfamily: gliadin

Query Match	21.9%	Score 75:	DB 2:	Length 411:
Best Local Similarity	37.0%	Pred.	No. 0.65:	
Matches 17:	Conservative	8:	Mismatches 17:	Indels 4:
				Gaps 3

QY 3 NPCAORCLQSCQQEPDDLKQKACESRCTKLEYDPRCVYDTG-ATNQ 47
 | | | | | | | | | |
Db 216 NMQNSCLNTCQQQQQPVGQ--CGSSCDQ--SQPQCATTSGSAYNQG 258

RESULT	5
T31863	

C:/date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:/Accession: T31863

submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid C54F6
A:Reference number: Z21094

A:Accession: AF011703
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-221 <GB>
 A:Cross-references: EMBL:AF016447; PDB:AAB65937.1; GSPDR:GN00023; CESP:C54F6.6
 A:Experimental source: strain Bristol N2; clone C54F6
 C:Genetics:
 A:Gene: CESP:C54F6.6
 A:Map position: 5
 A:Introns: 53/1; 85/2; 123/3; 177/1.

Query Match	20.4%	Score 70:	DB 2:	length 221:
Best Local Similarity	29.6%	Pred. NO. 1.4:		
Matches 16:	Conservative 7:	Mismatches 15:	Indels 16:	Gaps 2

```
Qy      3  NPCAQRCLDSCQQEPPDLKQKACESRCKLEY-----DPRCVYD  41
          ||| : | : : | ||| : ||| : ||||
Db     165  NNCAPKCFSSSPDKKDENVLK-CETKCAKLRTIMEEDDAKFSDFENCNACVYD  217
```

RESULT 6

hypothetical protein P3 - Mycoplasma hyorhinis
C:Species: Mycoplasma hyorhinis

R.Deng, G.; McIntosh, M.A.
J. Bacteriol. 176, 5929-5937, 1994

A1:Title: An amplifiable DNA region from the *Mycoplasma hyorhinis* genome
A:Reference number: Z18888; MUID:95014025

A;Accession: T18355
A;Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: DNA
A: Residues: 1-1187 <DEN>

A;Cross-references: EMBL:L11447; NID:g150156; PID:g664956; PIDN:AAA62228.1
C;Genetics:

A;Genetic code: SGC3

Query Match	20.1%;	Score 69;	DB 2;	Length 1187;
Best Local Similarity	28.6%;	Pred. No. 7.6;		

A:Cross-references: GB:L26081; NID:9799328; PIDN:AAA65938.1; PID:9436560
 C:Genetics:
 A:Gene: GDB:SEMA1
 A:Cross-references: GDB:283448
 C:Superfamily: semaphorin

Query Match 19.4%; Score 66.5; DB 2; Length 771;
 Best Local Similarity 35.2%; Pred. No. 9.8;
 Matches 19; Conservative 5; Mismatches 23; Indels 7; Gaps 2;

OY 6 AORCLOSCQEPDDLKQKACSRCTKLEYPRCVYDGNORHPGERTGRQ 59
 Db 510 AQLPLHRC-----DIYKCAECC--LADPYCAMDSACSRYFPJAKRTRRQ 556

RESULT 12
 T29699
 hypothetical protein F31A3.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
 C:Accession: T29699
 R:Murphy, J.; Le, T.-T.
 Submitted to the EMBL Data Library, May 1996
 A:Description: The sequence of C. elegans cosmid F31A3.

A:Reference number: Z20667
 A:Accession: T29699
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-242 <MUR>
 A:Cross-references: EMBL:U058742; PIDN:AAB36856.1; GSPDB:GN00028; CESP:F31A3.1
 A:Experimental source: strain Bristol N2; clone F31A3
 C:Genetics:
 A:Gene: CESP:F31A3.1
 A:Map position: X
 A:Introns: 18/3; 160/3
 C:Superfamily: ultra-high-sulfur keratin

Query Match 19.2%; Score 66; DB 2; Length 242;
 Best Local Similarity 43.3%; Pred. No. 4.1;
 Matches 13; Conservative 3; Mismatches 12; Indels 2; Gaps 1;

OY 2 ENPCARCLQSCQEPDDLKQKACSRCTK 31
 Db 175 EPQCQSCQCCVQCCVQCCSQMO--CASACTK 202

RESULT 13

T31887
 hypothetical protein C03A7.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T31887
 R:Greco, T.; Bradshaw, H.; Elliott, G.
 Submitted to the EMBL Data Library, July 1997
 A:Description: The sequence of C. elegans cosmid C03A7.
 A:Reference number: Z21096

A:Accession: T31887
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-388 <GRE>
 A:Cross-references: EMBL:AF016451; PIDN:AAB66001.1; GSPDB:GN00023; CESP:C03A7.4
 A:Experimental source: strain Bristol N2; clone C03A7
 C:Genetics:
 A:Gene: CESP:C03A7.4
 A:Map position: 5
 A:Introns: 75/3
 C:Superfamily: gliadin

Query Match 19.2%; Score 66; DB 2; Length 388;
 Best Local Similarity 43.8%; Pred. No. 6.2;

Matches 14; Conservative 4; Mismatches 8; Indels 6; Gaps 1;
 OY 5 CAORCLQSC-----QOEPDDLKQKACSRCT 30
 Db 317 CAPCEQSCCQCCVQCCQPPAACCTACSSCS 348

RESULT 14
 T31888
 hypothetical protein C03A7.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T31888
 R:Greco, T.; Bradshaw, H.; Elliott, G.
 Submitted to the EMBL Data Library, July 1997
 A:Description: The sequence of C. elegans cosmid C03A7.
 A:Reference number: Z21096

A:Accession: T31888
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-388 <GRE>
 A:Cross-references: EMBL:AF016451; PIDN:AAB65996.1; GSPDB:GN00023; CESP:C03A7.7
 A:Experimental source: strain Bristol N2; clone C03A7
 C:Genetics:
 A:Gene: CESP:C03A7.7
 A:Map position: 5
 A:Introns: 75/3
 C:Superfamily: gliadin

Query Match 19.2%; Score 66; DB 2; Length 388;
 Best Local Similarity 43.8%; Pred. No. 6.2;
 Matches 14; Conservative 4; Mismatches 8; Indels 6; Gaps 1;

OY 5 CAORCLQSC-----QOEPDDLKQKACSRCT 30
 Db 317 CAPCEQSCCQCCVQCCQPPAACCTACSSCS 348

RESULT 15

S06398
 alpha-globulin type A precursor - upland cotton

N:Alternate names: seed storage protein
 C:Species: Gossypium hirsutum (upland cotton)
 C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
 C:Accession: S06398
 R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
 Plant Mol. Biol. 9, 533-546, 1987

A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.
 A:Reference number: S06398
 A:Accession: S06398
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-605 <CHL>
 C:Superfamily: glycinnin
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-605/Product: alpha-globulin type A #status predicted <MAT>

Query Match 19.2%; Score 66; DB 2; Length 605;
 Best Local Similarity 38.3%; Pred. No. 9.1;
 Matches 23; Conservative 4; Mismatches 17; Indels 16; Gaps 5;

OY 9 CLQSCQ-----QOEPDDLKQ-KACSRCTKLEYP---RCVYDGNORHP-PGER 54
 Db 102 CEQSCREQYRKQOQPP--KQFKCQRCQWQRPKRKQCCVKECRQYQEDPWKGR 159

Search completed: March 1, 2001, 15:52:47
 Job time: 571 sec